Phonetograms, Aerodynamic Measurements, Self-Evaluations, and Auditory Perceptual Ratings of Male-to-Female Transsexual Voice

*Eva B. Holmberg, †Jennifer Oates, †Georgia Dacakis, and †Cameron Grant, *Stockholm, Sweden, †Victoria, Australia*

Summary: Objectives. This exploratory study reports instrumental and subjective data for 25 male-to-female transsexual (M-F TS) individuals using their attempted female voice. The aim was to examine the usefulness of phonetograms and aerodynamic measures for voice assessment of this client group.

Study Design. Descriptive and correlational.

Methods. Phonetogram speech-range profiles (SRPs) were recorded for the M-F TS participants' attempted female voice. Transglottal air pressure and airflow were estimated from oral recordings. All recordings were made in typical- and loud-voice conditions. Relationships among acoustical and aerodynamic measurements, background data, self-evaluations, and auditory perceptual ratings were examined. M-F TS data were compared with male and female normative data.

Results. Agreement between naive and voice-expert listeners as well as intra- and interlistener reliability was good. Fundamental frequency (F_0) accounted for 41–49% of variation in gender ratings for the group, but individual exceptions were found. Background data did not account for female voice success. Perceptual ratings of strain and breathiness were low. No data indicated hyperfunctional vocal behavior. The aerodynamic data agreed with normative male high-pitch data. The speech sound pressure level (SPL) was higher than the female norms. Phonetogram speech-range data fell between male and female data.

Conclusions. The importance of speaking fundamental frequency (SFF) in perception of gender was confirmed. Instrumental and subjective data suggested that the use of low speech intensities and avoidance of vocal fry could help contribute to a successful female voice. Phonetograms were suggested to be useful for visual feedback and documentation of changes in voice therapy for M-F TS clients.

Key Words: Transsexual–Gender–Phonetogram–Speech-range profile–Fundamental frequency (F_0)–Vocal intensity–Transglottal air pressure–Glottal airflow–Self-evaluations–Auditory perceptual voice ratings.

INTRODUCTION

"Transsexualism is a complex problem of gender identity in which the individual feels that his or her anatomic gender is the opposite of his or her psychological gender."[1] Most (75%) transsexual (TS) clients are males wishing to be reassigned as females.[2,3] The transition process of changing one's gender presentation is complex and usually involves hormonal treatment and sex-reassignment surgery. In addition, as the voice is an important gender marker, acquiring a sex-appropriate voice is an imperative part of the transition toward gaining acceptance in the TS individual's new gender. Vocal pitch is a strong gender marker,[4] and male-to-female TS (M-F TS) individuals who are perceived as females generally have higher mean speaking fundamental frequency (SFF) than those perceived as males.[5,6] Thus, much focus has been on helping M-F TS clients achieve and maintain female pitch characteristics. Hormone supplements of estrogen have no known biological effect on the male larynx and do not help to raise the fundamental frequency (F_0).[5,7] Sometimes, surgical procedures are used to achieve a higher fundamental frequency (F_0). These procedures include cricothyroid approximation,[8] anterior commissure advancement, and endolaryngeal shortening of the vocal folds.[9] However, although surgery can assist in raising the F_0, it is not problem-free and seldom sufficient to create a totally female voice.[10,11] Most of the M-F TS clients do not undergo pitch-raising surgery.

For most M-F TS clients, voice therapy is essential to bring the voice closer to a female voice, and much of the focus is on increasing SFF toward a female range. Oates and Dacakis[4] reported SFF for adult female (non-TS) Australian speakers to be between 145 and 275 Hz, with mean SFF values ranging between 196 and 224 Hz. Studies of M-F TS individuals have shown that, to be perceived as female, SFF needs to be between 155 and 160 Hz.[12,13] Gelfer and Schofield[5] showed that, for speakers who were perceived as women, SFF was between 164 and 199 Hz. However, the pitch target has to be carefully set for each M-F TS client,[4] and voice therapy needs to be individually designed.[14] In addition, although SFF is important for gender association, M-F TS individuals' own satisfaction with their voices is not necessarily related to their SFFs.[15] Voice features other than mean SFF, such as intonation pattern, articulation, formant patterns, and manner of speaking, are also gender markers.[1,13,16,17]

Accepted for publication February 17, 2009.

Different aspects of this research were presented at:

(1) The 7th Pan European Voice Congress (PEVoC07), Groningen, The Netherlands, August 29 to September 1, 2007.

(2) "Reflecting Connections," 2nd Conference hosted by the New Zealand Speech-Language Therapists Association and Speech Pathology Australia, Auckland, New Zealand, May 25–29, 2008.

From the *Clinical Science, Intervention, and Technology—CLINTEC, Division of Logopedics and Phoniatrics, Karolinska Institute, Stockholm, Sweden; and the †School of Human Communication Sciences, La Trobe University, Victoria, Australia.

Address correspondence and reprint requests to Eva B. Holmberg, Division of Logopedics and Phoniatrics, B09, Karolinska University Hospital, Huddinge, Stockholm, SE 141-86, Sweden. E-mail: eva.holmberg@ki.se

Journal of Voice, Vol. 24, No. 5, pp. 511-522

0892-1997/$36.00

doi:10.1016/j.jvoice.2009.02.002

Gelfer and Schofield[5] showed that, in addition to a higher SFF, M-F TS individuals who were perceived as females had a higher upper *SFF limit* than those perceived as males. One way to visualize and measure F_0 and sound pressure level (SPL) limits of speech is by phonetogram recordings. A phonetogram is an acoustic two-dimensional display of the voice in an SPL–F_0 coordinate system.[18–21] A third dimension, the color intensity of the registration, reflects how often all tones in the SPL–F_0 coordinate system are used. Phonetogram recordings have been used for illustration of differences between trained and untrained voices,[22] for changes pre- and post-voice therapy[23,24] and as a feedback system for singers.[25] In the present study, these phonetogram features were considered to have potentials for studies of TS voice.

Acoustical differences between male and female voices are mostly related to laryngeal structural differences and gender-dependent differences in voice aerodynamics. Transglottal air pressure is higher for males than for females and accompanied by higher vocal fold closing velocity.[26] As there is a strong positive relationship between these parameters and the SPL, males have generally louder voices than females. Male–female differences in glottal function also contribute to differences in voice quality.[27] Female voice is commonly produced with a posterior glottal opening between the arytenoids, a “chink” through which unmodulated airflow escapes.[28] The unmodulated airflow contributes to a steeper slope of the source spectrum with less harmonic energy in the high-frequency area, and the glottal airflow can lead to a higher degree of perceived breathiness in the female voice.[27] A somewhat breathy voice quality is commonly one of the goals of voice therapy for M-F TS clients.[4]

Gorham-Rowan and Morris[29] used flow inverse filtering to study glottal waveform differences between M-F TS speakers’ male and female voices. Their results showed that vocal fold closing velocity (as measured by the maximum flow declination rate [MFDR] of the glottal waveform) increased when the M-F TS speakers used their high-pitched female voices in comparison with their low-pitched male voices. Their results agree with inverse-filtering results of pitch change for non-TS males.[30] Increased vocal fold tension in combination with high vocal fold closing velocities is also typically found in hyperfunctional voice production.[31,32] Thus, production of a female voice with a male voice organ could be a potential risk for vocal fatigue or trauma to the folds and result in a perceptually strained voice quality.[33]

Although M-F TS individuals frequently undergo voice therapy to develop voice patterns that are close to those of biological females, there is a paucity of instrumental data on vocal function in TSs as compared with non-TS males and females. In the absence of such data, important aspects of glottal functioning for M-F TS clients’ modified and “new” voices are not known. Instrumental (“objective”) data are needed for research evidence regarding the outcomes of voice therapy and the factors that determine the success of voice therapy, and for making the underlying rationale for voice therapy methods for TS clients clear.[34]

The present study examined relationships between phonetograms, aerodynamic glottal data, background data, self-evaluations, and auditory perceptual ratings for a group of M-F TS clients. The study was exploratory and aimed to examine the usefulness of measurements of phonetogram and average glottal air pressure and airflow rate in the assessments of M-F TS voice.

METHODS

The study was conducted at the School of Human Communication Sciences at La Trobe University in Victoria, Australia. Ethical approval was obtained from the La Trobe University, Faculty of Health Sciences Human Ethics Committee (FHEC) before commencement of the study (Approval no.: FHEC 07/05). Each participant signed an informed consent form after agreeing to participate in the study.

Participants

Twenty-five Australian M-F TS volunteered to participate in the study. They were recruited from the Voice Clinic at the School of Human Communication Sciences, La Trobe University, and from the caseloads of three psychiatrists who provide most of the services to TS clients in Melbourne. None of the participants had undergone vocal fold surgery. The participants answered a questionnaire to provide information on their age, background, voice therapy, and progress in the gender-reassignment program. In terms of these parameters, the group was heterogeneous. Their ages ranged from 23.2 to 60.3 years with a mean of 44.8 years. Of the 25 participants, 22 had attended a gender-reassignment program. Among the three who had not attended a program, two were in the beginning of their gender-change process, whereas one lived full time as a woman but had never applied to a gender-reassignment program. The participants’ time in the program varied widely, ranging from 9 to 78 months, with a mean of 40.4 months. Fourteen participants were still in the program. Twenty of the 25 participants lived full time as women. The period for which they were full-time women varied from 9 to 144 months, with a mean of 42.7 months. Twenty-three of the 25 participants had received some voice therapy, but the number of sessions varied from 3 to 36, with a mean of 11.7 sessions. All but four participants were nonsmokers. The background data for the participants are presented in Table 1.

Assessments

The assessments were made over a period of 3 months. Each participant was assessed once. One hour was reserved for each participant, although this time was not always needed in full.

Self-report questionnaire

Before the acoustic and aerodynamic recordings, the participants completed a self-report questionnaire. In addition to the questions providing background information (Table 1), there were 10 questions, some of which included subquestions for alternative answers. The questions dealt with vocal health, the participant’s satisfaction with her voice, and whether she thought others perceived her as a woman. The questions were answered on horizontal 100-mm visual analogue scales (VAS)

TABLE 1.
Demographic Data for the M-F TS Group

	Age (y)	Time in Gender Program (mo)	Time Living as Woman (mo)	Number of Voice Therapy Sessions
	N = 25	N = 20	N = 20	N = 23
Mean	48.8	40	43	12
SD	11.3	21	33	9
Minimum	22.2	9	9	3
Maximum	60.3	78	144	36

Note: Gender program: N = 22, but two participants did not specify time period.

with "never/not at all satisfied" as the minimum (0 mm) and "all of the time/completely satisfied" as the maximum (100 mm).

Recordings

All recordings took place in a quiet, but not sound-treated, room. Before the recording, the participant was asked to have a drink of water to avoid sensations of a dry throat. The experimenter first briefly described the recording session and then gave detailed instructions before each recording set.

During the recordings, the participant was seated on a desk chair with a relaxed but upright posture for good breath support. A small microphone (AKG 420; Vienna, Austria) was used and fixed on a headset. The distance from the microphone to the corner of the mouth was adjusted for each participant to be 5 cm.

The experimenter monitored the signals on the laptop computer screen, but to encourage natural performance, the screen was located in a position so that the participant could not watch the signals during the recordings. The recording session began with the phonetogram recordings followed by the aerodynamic recordings. For all tasks, the participants were asked to use their female voices.

Recordings of phonetogram speech-range profiles.

Phonetogram recordings of speech-range profiles (SRPs) were made with the interactive computer program *Phog* (Saven Hitech AB, Stockholm, Sweden). The registration time for phonation was set to 25 milliseconds as recommended in the *Phog* instructions (Version 2.0, Real-time Phonetograph).

Calibration. The phonetogram system was calibrated at the beginning of each recording day following the calibration procedure recommended in the *Phog* instructions. A 1-kHz sinusoidal tone at 80 dB was generated by the *Phog* system and played through a loudspeaker (Fostex 6301B; Fostex International, Akishima, Tokyo, Japan). The microphone was connected to an analog-to-digital interface box with a Bullet 33 DSP card (Nyvalla; Communication Automation Corp., West Chester, PA, USA). The microphone level on the DSP audio interface box was adjusted to read 80 dB when the sound-level meter (RadioShack Corp., Fort Worth, TX, USA) was set on the linear scale and held 5 cm from the microphone on the headset (the *Phog* system automatically compensated the 5-cm distance for a distance of 30 cm).

Recording tasks. In all recording tasks, the participants used their attempted female voice.

The first task was to read a standard passage, "The Rainbow"[35] aloud in their typical voice. The text was presented on a laminated A4 poster and handheld by the participant. Before the recording, the participant read through the text to herself. It was explained to her that reading corrections did not matter.

The second task was a monologue. The participant was asked to talk about a topic of her choice in her typical voice for approximately 1 minute. If needed, topics were suggested. The recording was stopped between the reading and the monologue tasks, but the monologue phonetogram was superimposed on the reading phonetogram.

For the third and fourth tasks, the participant repeated the reading and monologue tasks in a loud voice. The instruction was to raise the voice as if talking to a group of people approximately 5 m away or to an audience. An SPL increase of 5–8 dB was the target, but not all participants managed to increase intensity by this amount in their intended female pitch. In these cases, the recordings were made of the participant's best possible attempt to produce loud voice. The recordings for typical and loud voices were saved in separate files.

Aerodynamic recordings of intraoral air pressure and oral airflow rate.

The F-J Electronics *Aerophone II* Software for Windows system (version June 14, 2005; F-J Electronics, Vedbaek, Denmark) was used for the aerodynamic recordings. For productions of strings of repeated /pæ/ syllables (pæpæpæpæpæ), simultaneous recordings were made of peak intraoral air pressure for /p/ (cm H_2O [centimeter water pillar]) and average intraoral airflow rate for /æ/ (L/s [litres per second]).[26,36–38] Each string consisted of five to seven syllables and was repeated at least three times. The aerodynamic signals, along with matching acoustic signals, were recorded on a Dell Latitude D610 laptop computer (Dell, Round Rock, TX, USA). *Intraoral air pressure* for the /p/ occlusions was captured with the use of a thin (inner diameter: 2.5 mm) silicone rubber catheter with one end passed between the participant's lips into the oral cavity approximately 1–3 cm behind the incisors, and the other end passed through the flow facemask and connected to a pressure transducer with a range of 0–30 cm H_2O. Average *oral airflow* rate for the /æ/ vowel was captured with the Aerophone mask and its differential pressure transducer system.[39]

Calibrations. Air pressure (cm H_2O) was calibrated with the use of a U-tube water manometer system. A calibration level of 10 cm H_2O was generated with a syringe and recorded along with a zero-level. Pressure calibration was done before the first recording session and then checked at each of the following four recording days, after which it was spot-checked throughout the study. The system was steady and did not have to be reset during the study period. *Airflow* (L/s) was calibrated before each participant's recording using the Aerophone II calibration procedures. The system was temperature sensitive and a high-quality room thermometer, placed close to the recording setup,

was read off before each participant's recording. Temperature was reset in the calibration system, when necessary.

Recording tasks. During the recordings, the participant was instructed to produce the /pæ/ syllables in each string linked together and in a monotonous and smooth fashion, a production that is needed to allow for estimates of glottal measures from the oral signals. The task was practiced before the recordings and repeated until there were three or more strings with signals that met the requirements for analysis.[26] The vowel quality varied somewhat among the participants, and productions with vowels approaching /a/ were accepted.

The participants produced the syllable strings first in *typical* and then in *loud* voice. To facilitate a natural production, the intensity and fundamental frequency (F_0) levels were not predefined or matched to the levels used in the SRP recordings. Typical- and loud-voice productions were saved in separate files.

Acoustic recordings for perceptual evaluations and F_0 analysis. Acoustic recordings of the whole recording sessions were made with the use of a Marantz Solid State PMD671 Flash recorder using the internal microphone (Marantz, D&M Professional, Itasea, IL, USA). Parts of these recordings were used for auditory perceptual analysis (the reading passage read in typical voice) and for F_0 analysis of the aerodynamic data (/pæ/ syllable strings in typical and loud voices). *Calibration* for this acoustic signal was the same as for the phonetogram recordings.

Data analyses

Speech-range profiles. Measurements of the SRPs in typical and loud voices were made with the use of the data analysis procedures included in the *Phog* program. The measurements included: *area* (semitones × SPL) of the SRP, the *lowest and highest* (minimum and maximum) F_0 (Hz) and SPL (dB) of the SRP, *mean values and standard deviations* (SDs) for F_0 (Hz) *and SPL* (dB). A provided measure of Leq, equivalent SPL, was highly correlated with SPL ($r = 0.99$), and thus, excluded from the data set.

Figure 1 presents an example of a phonetogram SRP. The measurement points for analyses of the *lowest and highest F_0 and SPL* are indicated by arrows. Single registrations at low frequencies reflecting vocal fry[40] were excluded from the analyses. Single registrations at high frequencies were also excluded as atypical for the speech. The excluded registrations are indicated in the Figure 1.

Air pressure, airflow, sound pressure level, and F_0. The Aerophone data analysis system was used for analyses of intraoral air pressure and oral airflow for the /pae/ syllable strings for typical- and loud-voice conditions separately. The number of analyzed syllables per loudness condition was approximately 10, with a minimum of 5, dependent on the quality of the signals.

Intraoral air pressure for the /p/ occlusions and *oral airflow* rate for the vowels in the /pæ/ syllable strings were used for estimation of *transglottal air pressure* (cm H_2O) and *glottal airflow* (L/s) for the vowels with a commonly used procedure.[41] Transglottal air pressure (ie, the pressure drop across the glottis) for the vowel was interpolated and averaged from peak intraoral pressures for the surrounding /p/ occlusions. Glottal airflow was extracted from the oral flow at a mid-vowel portion. SPL (dB) for the vowel was measured at a mid-vowel portion along with flow data.

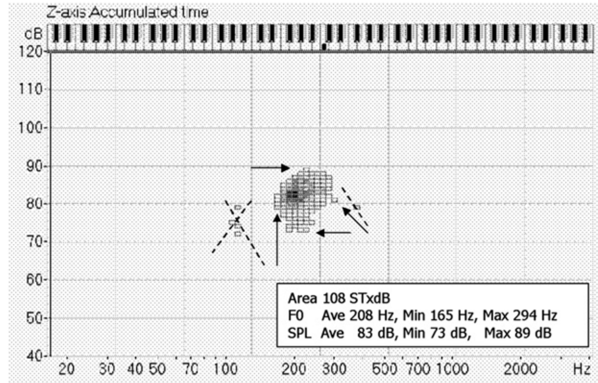

FIGURE 1. Speech phonetogram in typical voice for an M-F TS participant. Excluded registrations in low and high pitch (*dashed lines*) and measurement points for minimum and maximum F_0 and SPL (*arrows*) are indicated in the figure. All connected registrations were included in the measures. Values for area, F_0, and SPL are included.

F_0 (Hz) analysis for the syllable strings was completed separately with the use of the *Multi-Speech Main Program* (KayPentax, Lincoln Park, NJ, USA). In each of the condition of typical and loud voices, one string for which pressure, flow, and SPL values were consistent over all syllables and representative for the loudness condition was selected for the F_0 analysis. The selected syllable strings were extracted from the acoustic recordings with the use of the *Audacity program*: version 1.2.6 (http://audacity.sourceforge.net).

Auditory perceptual evaluations

Auditory perceptual evaluations were performed after completion of the participant recordings.

Procedures. Two listener groups provided the auditory perceptual ratings: (1) 20 naive listeners, and (2) two experienced Australian speech-language pathologists (SLPs) with voice as their area of expertise. All listeners were speakers of Australian English. The ratings were made on horizontal 100-mm VAS and measured in millimeters (mm).

1. The group of *naive listeners* (N = 20) consisted of 10 women and 10 men. The listeners were volunteers recruited from the academic staff of the Department of Statistics and Mathematics at La Trobe University and from the teaching staff of an elementary school close to the university. Their ages ranged between 24 and 58 years with a mean of 40 years (women—mean age: 41 years; men—mean age: 40 years). None had professional voice training or previous experience of voice evaluation. The naive listeners were told that the study was on voice

characteristics for adult speakers, but they were not given any information about the participants.

The recorded *listening material* for the naive listeners consisted of 65 readings of the Rainbow Passage in typical voice: 25 M-F TS participants' readings (the same recordings as used for the phonetogram speech-range analysis) and those by 12 non-TS men and 12 non-TS women in the same age range as the TS participants. The non-TS volunteers were recruited from associates of the researchers. They were all recorded using the same flash recorder that was used for the TS participants' recordings. The recordings were made in a quiet room at the School of Human Communication Sciences at La Trobe University or at the participant's home or workplace. All recordings were made by one of the investigators.

A total of 16 recordings were duplicated: eight of M-F TS participants, four of non-TS males, and four of non-TS females. All recordings were randomized and copied to CDs.

The listeners completed the rating task in a quiet room either at the School of Human Communication Sciences at La Trobe University or at their workplace. One of the investigators (J.O.) was present at all rating sessions and provided standard instructions to the listeners. The CD recordings were played to the listeners on a laptop or desktop personal computer (PC) with external speakers. The playback level was adjusted to be comfortable for all listeners. The listeners provided ratings of one parameter, "*gender*," on a rating form containing one VAS, for each speaker. The endpoints of the scale were marked "very male" (0) and "very female" (100). Each voice sample was played back once.

2. Two experienced Australian *SLPs* with voice as their area of expertise rated the 25 TS randomized voice recordings in consensus. They rated three parameters on VAS: *gender* (0 mm: very male—100 mm: very female—in the same way as the naive raters), *strain* (0 mm: none and 100 mm: severe), and *breathiness* (0 mm: none and 100 mm: severe). The 25 TS recordings were played back through a laptop computer with an external speaker. The playback level was adjusted so that it was comfortable for both SLPs. The SLP raters were provided with written instructions and rating sheets. They were informed that the recordings were of M-F TS individuals. They were permitted to listen to each voice sample several times if they wished.

Statistical analyses

For all statistical analyses, the type 1 error rate was set at $\alpha = 0.05$ in the interest of avoiding type 2 errors in this exploratory study. No comparative analyses were made between the two parameter sets with different speech material (phonetograms: reading/conversation; aerodynamic recordings: syllable repetition). Apart from descriptive summary statistics, the primary goals of the statistical analyses were as follows:

1. To determine the extent to which acoustic and aerodynamic characteristics, background data, self-evaluations, and auditory perceptual ratings of the M-F TS voice were systematically related to one another. Pairwise correlations were performed and r values calculated between all possible pairs. For correlation results with $r \geq 0.50$, regression analyses were carried out, and P values were listed in the results.
2. To determine whether there were significant differences in acoustic and aerodynamic voice characteristics between (1) the TS participants' typical and loud voices, (2) the TS participants and non-TS males, and (3) the TS participants and non-TS females. Independent or dependent t tests (two-tailed) were used as appropriate for these comparisons.

RESULTS

Self-evaluations

The participants' self-evaluations of vocal health and status are presented in Table 2.

As seen from the large SD and range values in Table 2, the M-F TS group was heterogeneous in terms of its self-rated voice experiences and evaluations. However, with the large variation in mind, mean values for questions on functional voice problems, such as hoarseness (questions 1 and 2a, b, c, d), and on questions related to vocal fatigue (questions 3, 4, and 5) were relatively low. Means for rated content with voice and pitch (questions 6 and 7) were higher than 50 mm (midscale) as were the means for ratings of how others perceived them as women (questions 8 and 9), all reflecting relatively

TABLE 2.
Summary Statistics for the M-F TS Participants' (N = 25) Self-Evaluations (mm) on a 100-mm VAS

Question	Mean	SD	Minimum	Maximum
1	38	22	0	66
2a	43	26	0	79
2b	45	26	0	80
2c	40	25	0	79
2d	31	25	0	76
3	50	25	9	90
4	35	23	0	87
5	31	22	0	71
6	58	27	1	97
7	62	24	5	100
8	53	31	6	100
9	66	30	5	100
10	51	34	0	100

Notes: Questions: (1) Is your voice generally croaky, hoarse, or husky? (2) Does your voice become croaky, hoarse, or husky in the following situations: (a) after prolonged use or loud talking in an everyday environment?; (b) after prolonged use or loud talking in an environment with background noise?; (c) after a late night?; (d) after attempts to increase your pitch? (3) Do you often clear your throat? (4) Do you experience dry and/or sore throat? (5) Does your voice get tired when you speak? (6) Are you content with your voice? (7) Are you content with the pitch of your voice? (8) "When I speak on the phone I am perceived as a woman." (9) "When I speak in social gatherings (e.g. café, hotel) I am perceived as a woman." (10) "I worry that my voice will expose my biological gender."

TABLE 3.
Results in mm for Gender Ratings on 100-mm VAS, With Endpoints Marked "Very Male" (0) and "Very Female" (100)

	Naive*			SLP†
	M-F TS	Non-TS F	Non-TS M	M-F TS
Mean	45.0	83.0	12.0	41.0
SD	19.2	3.3	6.9	14.0
Minimum	0.0	43.0	0.0	11.0
Maximum	100.0	100.0	63.0	69.0

*Naive listeners' (N = 20) ratings of M-F TS participants (N = 25), non-TS females (N = 10), and non-TS males (N = 10).
†SLP (N = 2) consensus ratings of M-F TS participants (N = 25).

positive experiences. At the same time, the ratings indicated worry about exposure of their biological gender (question 10).

Auditory perceptual evaluations

Gender ratings. Table 3 presents the results of gender ratings: (1) ratings made by the group of naive listeners (ratings of M-F TS participants, non-TS females, and non-TS males); (2) consensus ratings made by the voice-expert SLPs (ratings of M-F TS participants).

Intra- and interlistener reliability in ratings of gender. Intra- and interlistener reliability for the group of naive listeners (N = 20) was good: the mean intralistener reliability yielded an Intraclass Correlation Coefficient (ICC) of 0.843, and interlistener reliability yielded an ICC of 0.846.

Correlation analysis was performed between the naive listeners' ratings and the two SLPs' consensus ratings of gender. The results suggested good agreement between the naive and expert listeners ($r = 0.82$, $P < 0.001$).

Figure 2 presents a stylized illustration of the naive listeners' (N = 20) ratings of the male, female, and M-F TS speakers. The figure illustrates the ranges for male and female voices, shown as horizontal rectangular bars; male and female mean values, indicated with perpendicular lines; and range and mean for the M-F TS voices.

The mean value for non-TS male voice (Table 3) was 12 mm, and for non-TS female voice, it was 83 mm on the male–female 100-mm VAS. As illustrated in Figure 2, there was a male–female overlap in the gender ratings between 43 mm (lowest rating for non-TS female voice) and 63 mm (highest rating for non-TS male voice). Mean values for ratings of M-F TS voice (naive ratings: 45 mm; SLP ratings: 41 mm) fell in this gender-ambiguous area, close to the lowest ratings for non-TS female voice. However, the variation in rated gender of the TS group was large. The naive listeners' ratings ranged over the whole VAS (0–100 mm) as seen in Figure 2. The SLPs' gender ratings ranged from 11 to 69 mm on the VAS.

Gender ratings versus speaking fundamental frequency (F_0). Figures 3 and 4 (naive ratings and SLP ratings, respectively) present scatterplots of rated gender versus SFF for the M-F TS in typical voice. SFF accounted for 41% (naive: $r = 0.64$, $P = 0.001$) and 49% (SLP: $r = 0.70$, $P < 0.001$) of the variation in gender ratings. However, as indicated with arrows in Figure 3, some voices deviated from the group trend. Two of those received relatively high female ratings (79 and 82 mm on the 100-mm VAS) despite their relatively low SFF (150 and 135 Hz). Two other voices received low female ratings (34 and 38 mm) despite relatively high SFF (both 165 Hz).

Ratings of strain and breathiness. The SLPs' ratings of *strain* and *breathiness* in the M-F TS voices resulted in low values on the 100-mm VAS: *strain*—mean: 7, SD: 6, minimum: 0, and maximum: 21; *breathiness*—mean: 26, SD: 12, minimum: 10, and maximum: 54.

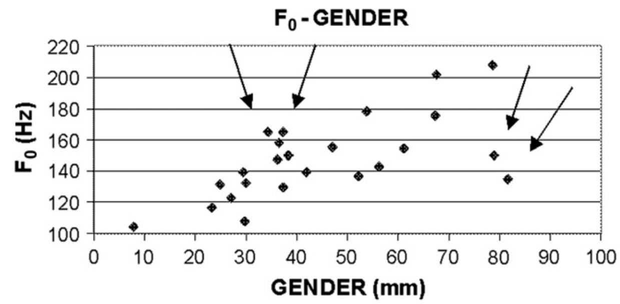

FIGURE 3. Scatterplots of gender versus F_0 for ratings made of the M-F participants (N = 25) by the naive listeners (N = 20). Individual data deviating from the general trend are indicated with *arrows*.

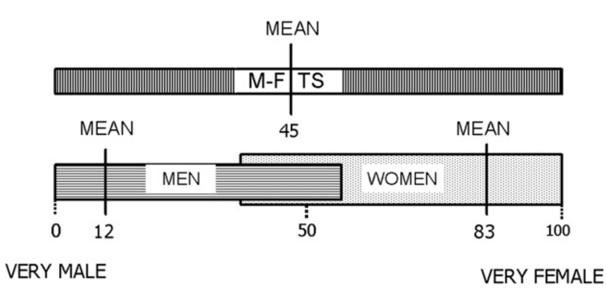

FIGURE 2. Stylized illustration of the naive listeners' (N = 20) ratings of the male (N = 10), female (N = 10), and M-F TS (N = 25) speakers. Ranges are shown as *horizontal rectangular bars*. Mean values are indicated with *perpendicular lines*.

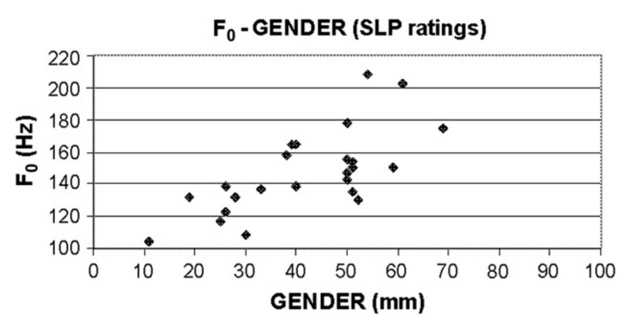

FIGURE 4. Scatterplots of gender versus F_0 for ratings made of the M-F participants (N = 25) by the SPL-expert listeners (N = 2 in consensus).

TABLE 4.
Summary Statistics for Phonetogram SRPs for the M-F TS Participants in: (1) Typical and (2) Loud Voices (N = 24)

	Mean	SD	Minimum	Maximum
Typical voice				
Speaking F_0 (Hz)	148.0	26.2	104	208
Speaking SPL (dB)	77.2	3.1	72	85
Area (ST × dB)	128.4	27.2	76	211
Highest frequency in area (Hz)	239.9	35.6	175	311
Lowest frequency in area (Hz)	110.3	25.0	73	175
Highest SPL in area (dB)	84.8	2.9	80	92
Lowest SPL in area (dB)	67.1	3.5	62	74
Loud voice				
Speaking F_0 (Hz)	161.2	26.0	114	210
Speaking SPL (dB)	80.9	3.4	74	91
Area (ST × dB)	144.0	30.6	86	211
Highest frequency in area (Hz)	256.0	44.5	185	370
Lowest frequency in area (Hz)	119.9	33.1	82	224
Highest SPL in area (dB)	89.1	3.3	83	100
Lowest SPL in area (dB)	69.5	4.2	61	77

Note: Phonetogram recordings: N = 25, missing data = 1.

Phonetogram speech-range profiles

Summary statistics for the M-F TS phonetogram SRP data in typical and loud voices are presented in Table 4 (N = 24). (Data are missing for one participant because of recording error.)

Dependent *t* tests were performed to examine differences in SRP measures between typical- and loud-voice productions. Mean SPL was significantly higher in loud voice ($P = 0.01$). Although F_0 was higher in loud voice and the speech area larger, these differences were not significant.

Pairwise correlation analyses were performed to investigate relationships between mean values, and maximum and minimum values of F_0 and SPL for the speech-profile areas in the typical- and loud-voice conditions: Relationships with $r \geq 0.50$ were found for the following parameter pairs: mean F_0 and minimum F_0 (typical: $r = 0.95$, $P < 0.001$; loud: $r = 0.72$, $P = 0.004$); mean F_0 and maximum F_0 (typical: $r = 0.71$, $P = 0.001$; loud: $r = 0.94$, $P < 0.001$); mean SPL and minimum SPL (typical: $r = 0.85$, $P < 0.001$; loud: $r = 0.81$, $P < 0.001$); mean SPL and maximum SPL (typical: $r = 0.88$, $P < 0.001$; loud: $r = 0.94$, $P < 0.001$).

Differences in sound pressure level and F_0 between the phonetogram and aerodynamic speech tasks

Because SPL and F_0 were not monitored during the recordings and the aerodynamic parameters could be expected to be related to SPL and F_0,[26,30] *t* tests were performed to examine SPL and F_0 differences between the phonetogram recordings and the aerodynamic recordings. In the *typical-voice* condition, mean F_0 was significantly higher in the aerodynamic recordings than in the phonetogram recordings ($P < 0.001$), whereas there was no significant difference in SPL. In the *loud-voice* condition, both F_0 and SPL were significantly higher in the aerodynamic recordings than in the phonetogram recordings ($P < 0.001$). Because of these differences in SPL and F_0 between the phonetogram and aerodynamic recordings, no statistical or qualitative comparisons were made between the two data sets.

Correlation results

Correlations ($r \geq 0.50$) with naive listeners' and speech-language pathologists' ratings of gender.

The correlation between naive listeners' versus SLPs' gender ratings was high ($r = 0.82$, $P < 0.001$). Table 5 presents pairwise correlations ($r \geq 0.50$) with rated gender. As seen in the table, both the SLP and naive gender ratings correlated with: SFF (in typical voice), content with pitch, content with voice, perceived as a woman on the phone, perceived as a woman at social gatherings, generally croaky voice (negative), and worry about gender exposure (negative).

Correlations ($r \geq 0.50$) with speaking fundamental frequency (F_0).

SFF in typical voice versus: *gender ratings* (*naive*: $r = 0.64$, $P = 0.001$; SLP: $r = 0.70$, $P < 0.001$); *content with pitch* ($r = 0.54$, $P = 0.006$); *perceived as a woman—on*

TABLE 5.
Correlations ($r \geq 0.50$) of Gender Ratings by the SLP (N = 2 consensus) and by the Naive Listeners (N = 20)

	SLP			Naive		
Gender Ratings—Versus	r	r^2	P	r	r^2	P
Speaking F_0 in typical voice	0.70	0.49	<0.001	0.64	0.41	<0.001
Content with pitch	0.54	0.29	0.006	0.55	0.22	0.033
Content with voice	0.50	0.20	0.017	0.50	0.21	0.014
Perceived as women on phone	0.77	0.59	<0.001	0.84	0.67	<0.001
Perceived as woman at gatherings	0.59	0.35	0.003	0.64	0.41	0.001
Generally croaky voice	−0.63	0.40	0.001	−0.72	0.53	0.001
Worry about exposure	−0.53	0.28	0.007	−0.68	0.50	<0.001

phone ($r = 0.59$, $P = 0.003$), *at social gatherings* ($r = 0.61$, $P = 0.002$).

Significant but weak correlations were found for *SFF in typical voice versus*: self-ratings of: generally *croaky voice* ($r = -0.46$, $P = 0.025$); *content with voice* ($r = 0.43$, $P = 0.037$); *worry about gender exposure* ($r = -0.47$, $P = 0.017$). Correlation between *SFF in typical versus SFF in loud voice*: $r = 0.86$, $P < 0.001$.

Correlations ($r \geq 0.50$) with speaking sound pressure level. Correlation between *SPL in typical versus SPL in loud voice*: $r = 0.88$, $P < 0.001$. There were no other significant correlations with speaking SPL.

Correlations ($r \geq 0.50$) with self-rated content with voice. Content with voice versus: *content with pitch* ($r = 0.80$, $P < 0.001$); *perceived as a woman on phone* ($r = 0.57$, $P = 0.005$); *perceived as a woman at social gatherings* ($r = 0.85$, $P < 0.001$); *croaky voice* ($r = -0.62$, $P = 0.001$); *worry about gender exposure* ($r = -0.66$, $P < 0.001$). Significant ($P < 0.05$) but weak correlations were found between content with voice and *SFF* ($r = 0.43$, $P < 0.037$) and *time living as woman* ($r = 0.42$, $P < 0.032$).

Additional correlations ($r \geq 0.50$). *Perceived as a woman on the phone and perceived as a woman at social gatherings* ($r = 0.61$, $P < 0.002$); *perceived as a woman on the phone and content with pitch* ($r = 0.57$, $P = 0.005$); *perceived as a woman at social gatherings and content with pitch* ($r = 0.85$, $P < 0.001$); *content with pitch and worry about gender exposure* ($r = -0.66$, $P = 0.0004$). A significant but weak correlation was found between content with pitch *versus* time living as a woman ($r = 0.43$, $P = 0.003$).

Parameters not correlated with $r \geq 0.50$ to any other parameter. No correlations with $r \geq 0.50$ were found for the following parameters: *time in gender program*, *time living as a woman*, *number of voice therapy sessions*; *frequency of throat clearing*; auditory perceptual ratings of *vocal strain and breathiness*.

Aerodynamic measurements

Summary statistics for estimated transglottal air pressure and glottal airflow along with SPL and F_0 for /pae/ syllable production in typical and loud voices are presented in Table 6.

Dependent *t* tests were performed to examine differences between typical and loud voices. All parameter values were significantly ($P < 0.05$) higher in loud voice with the exception of airflow, which did not differ between typical and loud voices. The results are presented in Table 7.

Relationships between transglottal air pressure, glottal airflow, F_0, and sound pressure level. Pairwise correlation analyses were performed to examine relationships between the transglottal air pressure, glottal airflow, F_0, and SPL in each of typical- and loud-voice syllable repetition. Significant ($P < 0.05$) relationships were found between transglottal air pressure and SPL (typical: $r = 0.82$, $P < 0.001$; loud: $r = 0.71$, $P < 0.001$). There were no further significant relationships.

TABLE 6.
Summary Statistics for Estimated Transglottal Air Pressure (P_{tg} [cm H_2O]), Glottal Airflow (V_g [L/s]), SPL (dB), and F_0 (Hz) for /pæ/ Syllable Productions in Typical and Loud Voice

	P_{tg}	V_g	SPL	F_0
Typical voice				
Mean	7.7	0.272	79.0	193.1
SD	1.9	0.117	3.4	29.6
Minimum	3.9	0.141	73.4	116.8
Maximum	11.0	0.532	83.3	244.7
Range	7.0	0.400	10.0	128.0
Loud voice				
Mean	10.6	0.298	86.0	210.2
SD	2.3	0.115	3.0	27.3
Minimum	6.2	0.134	81.6	133.3
Maximum	15.3	0.536	92.7	239.9
Range	9.0	0.400	11.0	107.0

TABLE 7.
Differences Between Typical and Loud Syllable Productions for SPL (dB), F_0 (Hz), Estimated Transglottal Air Pressure (P_{tg} [cm H_2O]), and Glottal Airflow (V_g [L/s])

	P values
SPL: typical – loud	<0.001
F_0: typical – loud	0.044
P_{tg}: typical – loud	<0.001
V_g: typical – loud	ns

Abbreviation: ns = nonsignificant.

DISCUSSION

An important part of TS individuals' gender change is the achievement of a gender-appropriate voice. Many M-F TS individuals receive voice therapy to help develop a female voice.[42,43] Auditory perceptual analysis of the voice is commonly used to evaluate the success of the voice therapy,[4] and much recent research focuses on developing instrumental ("objective") measures of voice. The instrumental data are used to relate perceptual manifestations of voice to quantitative measurements on the underlying vocal function. Quantitative data are also increasingly used in clinical voice evaluations,[41,44] and are needed to provide evidence for differences between normal and disordered voices and for monitoring change during voice therapy.[45–49] However, there is a paucity of instrumental data on M-F TS clients' vocal function. Comparisons between M-F TS voice data and normative male and female data should have strong clinical value in helping M-F TS clients achieve a female voice. The present study investigated the usefulness of two sets of instrumental measures for the evaluation of M-F TS female voice: noninvasive aerodynamic recordings for estimation of glottal air pressure and airflow, and phonetogram recordings of SRPs. The instrumental data were compared with the participants' background data and self-evaluations,

and with listeners' ratings of gender and voice quality. Several of the instrumental measures are shown to be significantly related to SPL,[26,27] and differences in measures between the M-F TS participants' (female) typical and loud voice were also studied.

The M-F TS participants formed a heterogeneous group both in terms of their background (age, time in gender program, time living as a woman, number of voice therapy sessions) and how successful they were in producing a female voice. As was illustrated in Figure 2, the M-F TS participants' voices received ratings across the entire VAS, ranging from "very male" to "very female." Correlation analyses showed that voice success was not related to background factors. For example, neither successful female voice nor F_0 were significantly related to the number of voice therapy sessions. In terms of successful female pitch, these results agree with those in a study by Dacakis,[43] who found no significant relationship between the achieved speaking F_0 and the number of therapy sessions. Lack of significant relationships between successful female voice and background data suggest that achievement of a female voice to a large extent depends on individual differences. Thus, the design of the voice therapy and its goals have to be individually set, as also pointed out previously by Dacakis.[14]

The heterogeneity of the M-F TS group was also reflected in the participants' self-ratings of vocal function (hoarseness, throat clearing, dry throat, and tired voice), content with voice and pitch, how they were perceived as women, and their worry about exposure of their biological gender. However, with the large variation in mind, the mean values of self-rated vocal dysfunction and vocal fatigue were relatively low, and the SLPs' low ratings of vocal strain were consistent with the participants' self-ratings. These results were somewhat surprising. The M-F TS clients in female high pitch must constantly violate the optimal use of their voice apparatus. It could be hypothesized that phonation with increased F_0 and attempt to produce a female voice would result in vocal hyperfunction[30] and cause functional problems and/or vocal fatigue. Indeed, current clinical guidelines recommend preventative measures, such as vocal hygiene education[50] and maximizing breath support,[42] in recognition of the potential for these problems. As 23 of the 25 participants had received voice therapy, the results of the current study support the relevance of this approach. However, for more complete understanding of the use of female voice and its impact on M-F TS clients' self-rated voice function, information on voice use and that from monitoring of vocal load is needed, and collection of such data is recommended for future studies.

Not surprisingly, parameters of the M-F TS participants' self-rated "content with pitch" and "perceived as a woman" were positively related to the perceptual gender ratings and successful female voice. Noteworthy was the significant negative relationship between gender rating and self-rated vocal fry ("croaky voice"). Voices with perceived vocal fry are likely to contain low-frequency energy,[51] and the results suggest that voice therapy should work on avoiding vocal fry to help achieve a successful female voice. The positive relationship between self-rated satisfaction with pitch and voice and "time living as a woman" may reflect that the M-F clients' voices became more female with time, but the results could also reflect that the clients had become accustomed to their voices or situation over time, or both.

Large intra- and interspeaker variation has been shown in measurements of average airflow,[52] which limits the usefulness of the measure. However, in combination with recordings of air pressure and SPL, average airflow measurements have been found to be clinically useful for the examination of changes in vocal behavior across vocal treatment.[39] The M-F TS female voice is anecdotally sometimes thought of being somewhat breathy. However, ratings of breathiness are based on subjective evaluations, and instrumental data on airflow are needed for the understanding of effects of potentially increased airflow. Increased glottal airflow would contribute to a steeper source spectrum slope with lesser high-frequency harmonic energy, similar to a female source spectrum.[27] Gorham-Rowan and Morris[29] studied M-F TS speakers' vocal function by means of flow inverse filtering. They found higher airflow parameter values for the speakers' female voices than for their male voices. However, the airflow values were not significantly related to gender ratings or successful female voice. Our results for average airflow agree in part with their data. Despite high average airflow, perceptual ratings of breathiness were low, and breathiness was not significantly related to gender ratings. From a clinical point of view, these results are positive, because a breathy voice quality *per se* is not a goal of voice therapy, but increased glottal airflow is merely used as a tool to help M-F TS clients in achieving a female voice quality.

Estimated transglottal pressure values for the M-F TS participants were close to those of non-TS males using increased pitch[30] despite the fact that F_0 was considerably higher and SPL was lower for the M-F TS participants than for the non-TS males. This finding may suggest that once in a high-pitch mode, pressure is not used to change F_0 in the same manner as between habitual and high pitch.[53–55]

In comparison with the M-F TS participants' typical voice, SPL was significantly higher in the loud voice in comparison with SPL in typical voice. Loud voice was produced with significantly increased transglottal air pressure. However, there was no significant change in glottal airflow. These results for loud voice agree with male and female non-TS data.[26]

Initially, we planned to monitor F_0 and SPL in the aerodynamic recordings to match values used in the phonetogram recordings. However, for some of the participants, the task of both keeping preset F_0 and SPL and producing the syllable strings in the smooth manner needed for reliable measurements was too difficult. To simplify the tasks, SPL and F_0 were not monitored, and mean F_0 was significantly higher for the syllable productions than in the speech tasks. Direct inferences were therefore not possible between the aerodynamic and phonetogram data sets.

Vocal pitch is a strong gender marker,[4] and a major focus of voice therapy for M-F TS clients is on increasing F_0 toward a female level. For a voice to be perceived as female, F_0 values around 155–165 Hz have been reported as lower limits.[5,12,13] In the present study, F_0 and rated gender were strongly correlated in both the naive listeners' and SLPs' ratings. The highest SLP rating was 69 mm, just above the naive listeners' highest

rating for a non-TS male voice (63 mm). However, the ratings differed in that the SLPs did not rate any M-F TS voice as all female (ie, 100 mm on the VAS) as did the naive listeners. This difference between the expert and naive ratings may depend on the differences in the rating procedures. The naive ratings were done individually by a relatively large group of listeners, whereas the expert ratings were done by two listeners in consensus. The SLPs also knew that the speakers were M-F TS persons, whereas the naive listeners did not have any information about the speakers. For the naive ratings, the M-F TS voices were also mixed with non-TS male and female voices, whereas the SLPs rated only the M-F TS voices. The different listening procedures were made to mimic realistic situations in terms of how naive listeners and clinicians meet TS individuals; the naive listeners in society among other men and women; the clinicians as voice clients with a known problem. Despite the differences in listening procedures, the high agreement on gender between the SLPs and naive listeners is positive, not only for research purposes, but also clinically for decisions of voice therapy goals.

Despite the strong relationship between F_0 and rated gender, the scatterplot data of the naive listeners' gender ratings versus F_0 (Figure 3) show exceptions from the group trend. Two of the participants with the highest gender ratings (79 and 82 mm) on the male–female VAS (100 mm) had relatively low F_0 (150 and 135 Hz)—below the frequency usually considered as the limit for a female sounding voice. In contrast, two other participants with higher F_0 (both 165 Hz) received low ratings (34 and 38 mm) on the male–female scale. In the background data, no consistent similarities or differences were found that could explain these individual results for gender ratings versus F_0. One of the two *successful* participants had not applied for or participated in any gender program and the other had completed the program; one of the two *unsuccessful* participants had not started her gender program, whereas the other one had completed it. In terms of voice therapy, one of the two participants with *unsuccessful* female voice had received four therapy sessions and the other received 41 sessions. Noteworthy was that neither of the participants with *successful* voice had received any voice therapy. These mixed findings agree with previous studies of M-F TS clients, which have shown that, apart from F_0, a successful female voice depends on a set of contributing factors.[5]

One parameter that separated these participants with successful and unsuccessful female voices was *mean SPL*, which was somewhat lower for the two successful participants (75 and 76 dB) in comparison with the two unsuccessful (79 and 80 dB) participants. In addition, values of *minimum* SPL, that is, the lowest SPL in the phonetogram SRPs were lower for the successful (63 dB for both participants) than for the unsuccessful participants (66 dB for both). The individual results suggest that the two successful participants used more low-intensity voice in their speech.

The individual SRP results can be compared with group results in Figure 5.

Figure 5 shows stylized SRPs drawn from group mean values for the M-F TS participants in typical voice in comparison with SRPs drawn from group mean values for men and women[56]

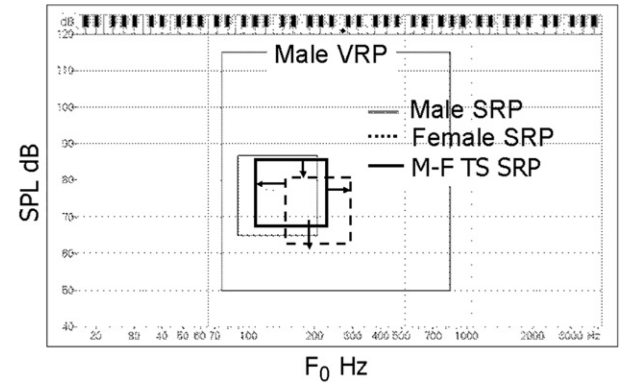

FIGURE 5. Stylized phonetogram SRP for non-TS males,[56] non-TS females (unpublished data), and the M-F TS participants inserted in a maximum voice range profile (VRP) for non-TS males.[56]

according to a model by Ma and Yui.[57] The speech profiles are inserted in a stylized maximum voice SRP for males.[58] The maximum range represents the maximum voice capacity for M-F TS clients who have not undergone laryngeal surgery. Speech and maximum range profiles are shown in a computer phonetogram display.

As illustrated in Figure 5, the mean M-F TS SRP area (128 ST*dB) fell between mean SRP areas for men (142 ST*dB) and for women (91 ST*dB). Maximum (240 Hz) and minimum (110 Hz) F_0 for the M-F TS participants fell between male (maximum: 198 Hz, minimum: 89 Hz) and female (maximum: 308 Hz, minimum: 162 Hz) values. M-F TS maximum SPL (85 dB) was very close to the maximum SPL for men (86 dB), and higher than the maximum SPL for women (80 dB). M-F TS minimum SPL (67 dB) was higher than those for both men (65 dB) and women (64 dB). The individual and group data suggest that lower SPL and increased use of low voice intensities may help contribute to a more successful female voice, and soft voice in female pitch should be practiced in voice therapy for M-F TS clients.

In addition to speech-range data, the phonetogram display provides visual feedback, which can facilitate the acquisition of independent control over pitch and loudness. Phonetograms have been found to be a valuable tool in voice training for singers[25] and have also been used for documentation of changes in voice therapy.[24] Phonetogram recordings should be a useful tool for visual feedback in voice therapy for M-F TS clients and for objective documentation of change in voice therapy.

CONCLUSIONS

The significance of fundamental frequency (F_0) for gender assessment was confirmed in this study. However, phonetogram SRP data suggested that the use of low speech intensities could also contribute to a successful female voice. No indications of vocal strain were found in self-evaluations, auditory perceptual evaluations, or in any of the instrumental data. It was suggested that measurements of vocal use and vocal load would be important for the evaluation of M-F TS vocal function in future studies. In addition, the combination of relatively high values of airflow and low values of perceived breathiness suggest that acoustic spectral

measurements could add to the understanding of M-F TS attempted female voice. Combined results suggested that phonetograms and aerodynamic measurements should add to the evaluation and documentation of voice therapy for M-F TS clients.

Acknowledgments

This study was supported by a Fellowship from the Institute for Advanced Studies, La Trobe University, and by a grant from the Swedish Voice Foundation.

We are grateful to Sheryl Mailing for assistance in the perceptual ratings, to Shane Erickson for assistance with statistical analyses, to Anna Nilsson for phonetogram recordings of the non-TS women, and to the TS clients and non-TS volunteers for their participation in the study. We also thank Britta Hammarberg and Maria Södersten for readings of a previous version of the manuscript.

REFERENCES

1. Oates J, Dacakis G. Speech pathology considerations in the management of transexualism—a review. *Br J Disord Comm*. 1983;18:139–151.
2. Van Kersten PJ, Gooren LJ, Megens JA. An epidemiological and demographic study of transsexuals in the Netherlands. *Arch Sex Behav*. 1996;25:589–600.
3. Van Borsel J, De Cuypere G, Van den Berghe H. Physical appearance and voice in male-to-female transsexuals. *J Voice*. 2001;15:570–575.
4. Oates J, Dacakis G. Voice change in transsexuals. *Venereology*. 1997;10: 178–187.
5. Gelfer M, Schofield KJ. Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as males. *J Voice*. 2000;14:549–556.
6. Söderpalm E, Larson A, Almquist SÅ. Evaluation of a consecutive group of transsexual individuals referred for vocal intervention in the west of Sweden. *Logoped Phoniatr Vocol*. 2004;29:18–30.
7. Schapira K, Davidson K, Brierley H. The assessment and management of transsexual problems. *Br J Hosp Med*. 1979;12:63–67.
8. Kanagalingam J, Georgalas C, Wood G, Ahluwalia S, Sandhu G, Cheesman A. Cricothyroid approximation and subluxation in 21 male-to-female transsexuals. *Laryngoscope*. 2005;115:611–618.
9. Gross M. Pitch-raising surgery in male-to-female transsexuals. *J Voice*. 1999;13:246–250.
10. Yang CY, Palmer AD, Murray KD, Meltzer TR, Cohen JI. Cricothyroid approximation to elevate vocal pitch in male-to-female transsexuals: results of surgery. *Ann Otol Rhinol Laryngol*. 2000;111:477–485.
11. Van Borsel J, Van Eynde E, De Cuypere G, Bonte K. Feminine after cricothyroid approximation? *J Voice*. 2008;22:379–384.
12. Spencer LE. Speech characteristics of male-to-female transsexuals: a perceptual and acoustic study. *Folia Phoniatr*. 1988;40:31–42.
13. Wolfe VI, Ratusnik DL, Smith FH, Northrop G. Intonation and fundamental frequency in male-to-female transsexuals. *J Speech Hear Disord*. 1990;55: 43–50.
14. Dacakis G. Long-term maintenance of fundamental frequency increases in male-to-female transsexuals. *J Voice*. 2000;14:549–556.
15. McNeill EJM, Wilson JA, Clark S, Deakin J. Perception of voice in the transgender client. *J Voice*. 2007;22:727–733.
16. Coleman R. Male and female voice quality and its relationship to vowel formant frequencies. *J Speech Hear Res*. 1971;14:565–577.
17. Coleman R. A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *J Speech Hear Res*. 1976;19:168–180.
18. Gramming P. *The phonetogram: an experimental and clinical study*. Malmö, Sweden: Department of Otolaryngology, Lund University; 1988 [doctoral thesis]1–162.
19. Pabon JP, Plomp R. Automatic phonetogram recording supplemented with acoustic voice quality parameters. *J Speech Hear Res*. 1988;31:710–722.
20. Pabon JPH. Objective acoustic voice-quality parameters in the computer phonetogram. *J Voice*. 1991;3:203–216.
21. Heylen L, Wuyts FL, Mertens F, De Bolt M, Van de Heyning PH. Normative voice range profiles of male and female professional voice users. *J Voice*. 2002;16:1–7.
22. Awan SN. Phonetographic profiles and F0-SPL characteristics of untrained versus trained voice groups. *J Voice*. 1991;5:41–50.
23. Speyer R, Wieneke GH, van Wijck-Warnaar I, Dejonchere PH. Effects of voice therapy on the voice range profiles of dysphonic patients. *J Voice*. 2003;17:544–556.
24. Holmberg EB, Ihre E, Sodersten M. Phonetogram as a tool in the voice clinic: changes across voice therapy for patients with vocal fatigue. *Logoped Phoniatr Vocol*. 2007;32:113–127.
25. Lamarche A, Ternström S, Hertegård S. Not just sound: supplementing the voice range profile with the singer's own perception of vocal challenges. Logoped Phoniatr Vocol. 2009;34:3–10.
26. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J Acoust Soc Am*. 1988;84:511–529.
27. Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am*. 1990;87: 820–857.
28. Södersten M, Lindestad P-Å, Hammarberg B. Vocal fold closure, perceived breathiness, and acoustic characteristics in normal adult speakers. In: Gauffin J, Hammarberg B, eds. *Vocal Fold Physiology—Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*. San Diego: Singular Publishing Group, Inc; 1991:217–224.
29. Gorham-Rowan M, Morris R. Aerodynamic analysis of male-to-female transgender voice. *J Voice*. 2006;20:251–262.
30. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch. *J Voice*. 1989;3:294–305.
31. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: an experimental framework and initial results. *J Speech Hear Res*. 1989;32:373–392.
32. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughan C. Phonatory function associated with hyperfunctionally related vocal fold lesions. *J Voice*. 1990;4:52–63.
33. Holmberg EB, Hillman RE, Hammarberg B, Södersten M, Doyle P. Efficacy of a behaviorally based voice therapy protocol for vocal nodules. *J Voice*. 2001;15:395–412.
34. Oates J. Evidence based practice. In: Adler RK, Hirsch S, Mordaunt M, eds. *Voice and Communication Therapy for the Transgender/Transsexual Client. A Comprehensive Clinical Guide*. San Diego: Plural Publishing; 2006:23–44.
35. Fairbanks G. *Voice and Articulation Drill Book* (2nd ed.). New York: Harper. 1960.
36. Netsell R. Speech physiology. In: Minifie FD, Hixon TJ, Williams F, eds. *Normal Aspects of Speech, Hearing and Language*. Englewood Cliffs, NJ: Prentice-Hall; 1973:211–234.
37. Smitheran JR, Hixon TJ. A clinical method for estimating laryngeal airway resistance during vowel production. *J Speech Hear Res*. 1981;46: 138–146.
38. Löfqvist A, Carlborg B, Kitzing P. Initial validation of an indirect measure of subglottal pressure during vowels. *J Acoust Soc Am*. 1982;72: 633–634.
39. Zeitels SM, Hillman RE, Desloge R, Mauri M, Doyle P. Phonosurgery in singers and performing artists: treatment outcomes, management theories, and future directions. *Ann Otol Rhinol Laryngol*. 2002;12(suppl 190):21–40.
40. Zraick RI, Nelson JL, Montague JC, Monoson PK. The effect of task on determination of maximum phonational frequency range. *J Voice*. 2000;14:154–160.
41. Hillman RE, Montgomery WW, Zeitels SM. Current diagnostics and office practice: appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders. *Curr Opin Otolaryngol Head Neck Surg*. 1997;5:172–175.
42. Becklund Fridenberg C. Working with male-to-female transgendered clients: clinical considerations. *Contemp Issues Commun Sci Disord*. 2002;29:43–58.

43. Dacakis G. The role of voice therapy in male-to-female transsexuals. *Curr Opin Otolaryngol Head Neck Surg*. 2002;10:173–177.
44. Mehta DD, Hillman RE. Voice assessment: updates on perceptual, acoustic aerodynamic, and endoscopic imagin methods. *Curr Opin Otolaryngol Head Neck Surg*. 2008;16:211–215.
45. Carding P. *Evaluating Voice Therapy: Measuring the Effectiveness of Treatment*. London and Philadelphia: PA Whurr Publishers Ltd; 2000.
46. Dollaghan CA. *The Handbook for Evidence-Based Practice in Communication Disorders*. Maryland: Brookes Publishing Co., 2007.
47. Reilly S, Oates J, Douglas J. *Evidence Based Practice in Speech Pathology.* Sweden: John Wiley and sons Ltd; 2003.
48. American Speech-Language-Hearing Association. Evidence-based practice in communication disorders: an introduction.http://www.asha.or/members/deskref-journals/deskref/default; 2004. [Technical report].
49. Ma EPM, Yiu EM-L, Verdolini Abbott K. Application of the ICF on voice disorders. *Semin Speech Lang*. 2007;28:343–350.
50. Adler RK, Mordaunt M, Hirsch S, eds. *Voice and Communication Therapy for the Transsexual/Transgender Patient: A Complete Clinical Guide*. San Diego: Plural Publishing Corporation; 2006.
51. Hammarberg B, Fritzell B, Gauffin J, Sundberg J, Wedin L. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol*. 1980;90: 441–541.
52. Holmberg EB, Hillman RE, Perkell JS, Gress C. Relationships between intra-speaker variation in aerodynamic measures of voice production and variations in SPL across repeated recordings. *J Speech Hear Res*. 1994;37:484–495.
53. Stevens KN. Physics of laryngeal behavior and larynx modes. *Phonetica*. 1977;34:264–279.
54. Sundberg J, Titze I, Scherer R. Phonatory control in male singing: a study of the effects of subglottal pressure, fundamental frequency and mode of phonation on the voice source. *J Voice*. 1993;7:15–29.
55. Holmberg EB, Perkell JS, Hillman RE, Gress C. Individual variation of voice. *Phonetica*. 1994;51:30–37.
56. Hallin AE, Berglund K, Holmberg EB, Södersten M. Voice range profiles for vocally healthy men; normal data and methodological issues. 27th IALP Congress Kgs. Lyngby, Denmark; August 5–9, 2007.
57. Ma EPM, Yiu EML. Multiparametric evaluation of dysphonic severity. *J Voice*. 2006;20:380–390.